Center Reflections

A monthly publication highlighting activities at the W.M. Keck Foundation Center for Molecular Structure

California State University Fullerton

Volume 1, Issue 5 May 1999

New Tools Available in the Cambridge Structural Database

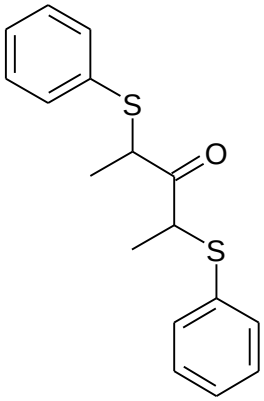
The Cambridge Structural Database (CSD) contains crystal structure information for over 190,000 small molecule organic and organometallic compounds. It is the only comprehensive collection of entries determined by X-ray diffraction and neutron diffraction techniques.

For each crystallographic entry, there are three distinct types of information stored, conveniently categorized by their "dimensionality".

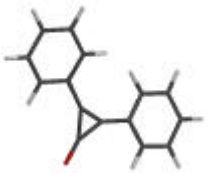
1D Information - These data fields incorporate all of the bibliographic material for a particular entry and summarize the structural and experimental information for the crystal structure. The text and numerical information includes the authors' names and the full journal reference, as well as the crystallographic cell dimensions and space group.

LAJCOO
N-Acetyl-alpha,beta-dehydro-O-methyltyrosyl-beta-alanine
antimorphine activity
C15 H18 N2 O5
A.A. Karapetyan, V.O. Topuzyan, Yu.T. Struchkov
Zh. Strukt. Khim., 33, 151-3, 1992

2D Information - A conventional chemical diagram of the molecule is stored in this information field. This is encoded as a chemical connection table comprising atom and bond properties. Atom properties include element symbol, number of connected non-hydrogen atoms, number of connected hydrogen atoms, and the net charge. Seven different bond types can be specified in bond properties.



3D Information - A 3D representation of the molecule can be generated from the information stored in this field. This data includes the atomic coordinates, the space group symmetry, the covalent radii, and the crystallographic symmetry established by using those radii. The 3D representation is matched with the 2D chemical structure.



Among the most exciting, fastest growing, and potentially useful areas of X-ray crystallography are studies of supramolecule structures, molecular recognition, solid state reactions, and crystal design. New tools introduced into the Cambridge software package allow users to view data in new ways, answering old questions and suggesting new ones. Specific intermolecular interactions can now be displayed on 6000 scatter plots. The location of the molecular centers in different space groups can be analyzed. Patterns characteristic of molecular interaction in the CSD can be identified, and explanations for the stereo chemical and electronic properties that lead to these interactions can be derived.

The latest release of the CSD and its new tools are now available from CMolS for use. To obtain subscription information and an account, please contact Katherine Kantardjieff or Brian Schick.

25 Years of the Protein Data Bank

If one picture is worth a thousand words, recent advances in X-ray crystallography methods are providing the equivalent of the *Encyclopedia Britannica*. Crystallographers are now churning out three-dimensional structures of proteins at the rate of four per day!

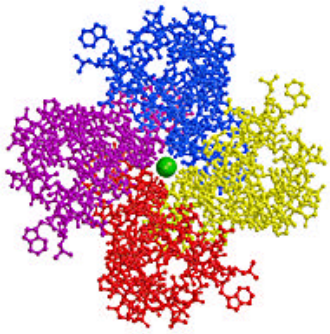
The Protein Data Bank (PDB) celebrated its 25th year of service to the structural biology community in 1998. The PDB is the single international repository for the processing and distribution of three-dimensional structure data of biological macromolecules determined experimentally by X-ray crystallography and NMR. From its inception in 1973, it took until 1992 before 1,000 depositions were made. There are now over 9,800 entries! The current rate of submission is about 50 entries per week, compared with about three per week in the early 1990s.

The PDB has been making changes to improve its services, such as developing tools to visualize, query and understand the entries in the bank. Further developments include a web-based deposition system, establishment of mirror sites, a new browser that includes automatic links to external data, and design and release of a standard structure factor file format similar to the CIF.

The award for the operation of the PDB in the United States is being transferred to the Research Collaboratory for Structural Bioinformatics (RCSB). Brookhaven National Laboratory (BNL) and RCSB have collaborated on a phased transition of responsibilities which began on October 1, 1998, and was anticipated to be completed on October 31, 1999. However the transition has proceeded ahead of schedule, allowing for an earlier complete transfer of responsibilities.

Operations of the PDB at BNL will cease on June 30, 1999. The BNL PDB website (http://pdb.pdb.bnl.gov/) will no longer be available after that date. The RCSB home page is at http://www.rcsb.org/. Access to the PDB is still free of charge to users but the future of this free access remains uncertain.

The structures in the PDB whose atomic coordinates have been determined by X-ray crystallography are included with the latest version of the Cambridge Structural Database at CMolS.



Industry User Fee for SWISS-PROT

Plagued by a funding crunch and inundated with new data, SWISS-PROT, a widely used amino acid database linked to the PDB, will soon start charging a fee to industrial users. SWISS-PROT contains sequences and other information on more than 70,000 proteins and is used by some 200,000 researchers in 100 countries, according to Amos Bairoch of the University of Geneva, who developed SWISS-PROT. Managers, however, have a backlog of about 150,000 computer-generated sequences from which to cull protein information. The budget for the database, at $3M, is only half of what it needs.

It's in the Bank

GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (Nucleic Acids Research 1998 Jan 1;26(1):1-7). There are approximately 2,570,000,000 bases in 3,525,000 sequence records as of April 1999. Sequences organisms include viruses, bacteria, yeast, and a number of plasmids and organelles. As an example, you may view the record for the neurofibromatosis gene. The complete release notes for the current version of GenBank are available. A new release is made every two months. GenBank is part of the International Nucleotide Sequence Database Collaboration, which is comprised of the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis. Find out more at http://www.ncbi.nlm.nih.gov/.

NIST Crystal Data

NIST Crystal Data contains chemical, physical, and crystallographic information useful to characterize more than 210,403 inorganic and organic crystalline materials. The data include the standard cell parameters, cell volume, space group number and symbol, the calculated density, and classification by chemical type, chemical formula, and chemical name. Each entry has an associated literature reference.

The database can be utilized as a practical analytical tool for compound identification because the lattice/formula combination uniquely characterizes a crystalline phase. The database is useful in conjunction with other data for materials design and properties prediction. The file includes reliable data across the entire range of solid state materials including inorganics, organics, minerals, intermetallics, metals, alloys, drugs, antibiotics, and pesticides. Comprehensive chemical, crystallographic, and identification search software is provided with the database.

The database is available in magnetic tape and CD-ROM formats. You may search it interactively using the Bruker software available at CMolS.

National Institute of Standards and Technology

http://www.nist.gov/srd/srd.htm

NIST DATA

Websites of Interest

- ***Medline search*** (www.nlm.nih.gov) - now open to the public for free, no registration required, courtesy of the National Library of Medicine, a branch of the NIH.

- ***Annals of Improbable Research*** (www.improb.com) - rare and well-done tidbits.

- ***NIST Online Reference Databases*** (www.nist.gov/srd/online.htm) - include a section on physical reference data such as physical constants, units and conversion factors; atomic and molecular spectroscopy; nuclear physics; and condensed matter physics.

- ***ESG Biology Hypertextbook*** (www.mit.edu/afs/athena.mit.edu/course/other/esgbio/www/index.html) - covers biology, chemistry, biochemistry and recombinant techniques.

- ***Crystallography 101*** (www-structure.llnl.gov/Xray/101index.html) - see Volume 1, Issue 3 of *Center Reflections*.

- ***The Impact of Electronic Publishing on the Academic Community*** (tiepac.portlandpress.co.uk) - experts from a wide variety of backgrounds discuss the plans for implementing electronic publishing in their specific subject areas. The online version of this book is fully searchable with links from the text to references and hot links to other websites and e-mail addresses.

Upcoming Events

May 22-26, 1999: **American Crystallographic Association** National Meeting, Buffalo, NY. http://nexus.hwi.buffalo.edu/ACA/ACA-Annual/Buffalo/Buffalo.html

June 26-28, 1999: **Computational Science Workshop**, San Diego State University. http://www.edcenter.sdsu.edu/training/workshop/events.html

July 8-10, 1999: **CSU Computational Chemistry Council** Summer Conference, Humboldt State University. http://www.humboldt.edu/~chemcomp/

August 4-13, 1999: **International Union of Crystallography** Meeting, Glasgow, Scotland. http://www.iucr.org/

August 22 - 26, 1999: **American Chemical Society** National Meeting, New Orleans, LA. http://www.acs.org/meetings/neworleans/welcome.htm

October 6-8, 1999: **Pacific Conference on Chemistry and Spectroscopy** and **35th ACS Western Regional Meeting**, Ontario Convention Center. The plenary lecture on October 6 will be given by Jerome Karle. Abstracts are due May 15. Abstract forms may be obtained from the ACS website. http://www.acs.org/meetings/abstract/abrqst.html.

October 26-29, 1999: **EDUCAUSE '99** "Celebrating New Beginnings", Long Beach Convention Center. Information technology conference. http://www.educause.edu/conference/e99/.

W.M. Keck Foundation Center for Molecular Structure

Department of Chemistry and Biochemistry

California State University Fullerton

800 N. State College Blvd.

Fullerton, CA 92831

http://zeppo.fullerton.edu:8080/~kkant/cmols2.html

Director: Dr. Katherine Kantardjieff

kkantardjieff@fullerton.edu

Staff Scientist: Dr. Brian Schick

bschick@doc.fullerton.edu